Virology **266,** 99–109 (2000)
doi:10.1006/viro.1999.0066, available online at http://www.idealibrary.com on IDEAL®

On the Domain Structure and the Polymerization State of the Sendai Virus P Protein

Nicolas Tarbouriech,* Joseph Curran,† Christine Ebel,‡ Rob W. H. Ruigrok,§ and Wilhelm P. Burmeister*,[1]

*ESRF, B.P. 220, 38043 Grenoble cedex 9, France; †Department of Microbiology, University of Geneva Medical School, CMU, 9 avenue de Champel, CH-1211 Geneva 4, Switzerland; ‡Institut de Biologie Structurale CEA-CNRS, 41 rue Jules Horowitz, 38027 Grenoble cedex 1, France; and §EMBL Grenoble Outstation, c/o ILL, B.P. 156, 38042 Grenoble cedex 9, France

Received June 8, 1999; returned to author for revision June 29, 1999; accepted October 25, 1999

The phosphoproteins (P) of paramyxoviruses and rhabdoviruses are cofactors of the viral polymerase (L) and chaperones of soluble nucleoprotein preventing its polymerization and nonspecific binding to cellular RNA. The primary sequences of six paramyxovirus P proteins were compared, and although there was virtually no sequence similarity, there were two regions with similar secondary structure predictions in the C-terminal part of P: the predicted multimerization domain and the X-protein, the sequence that binds to N in the N:RNA template. The C-terminal part of the Sendai virus P protein, the multimerization domain including the binding site for the polymerase, and the X-protein were expressed in *Escherichia coli.* All three polypeptides folded with secondary structures similar to those predicted. The C-terminal part of P is a very elongated molecule with most of its length encompassing the multimerization domain. Both the multimerization domain and the C-terminal part of P were found to form tetramers, whereas the X-protein was monomeric.

INTRODUCTION

Paramyxoviruses are enveloped animal viruses containing a nonsegmented negative-stranded RNA genome. Sendai virus (SeV), a prototype paramyxovirus, has served as a model for examining the mechanisms that regulate viral genome expression. The template for RNA synthesis is the helical genome (or antigenome) nucleocapsid in which each N protein subunit is associated with precisely six nucleotides (Egelman *et al.,* 1989). Only viral genomes that are $6n + 0$ nucleotides in overall length are found naturally or are fully competent for replication as minigenomes (Calain and Roux, 1993). This restriction is thought to operate during the initiation of genome synthesis because the viral promoter sequences at the 3′ ends of the templates are recognized in the context of the N-protein subunits that impose a hexamer phase (for a review, see Kolakofsky *et al.,* 1998).

The SeV polymerase is a complex of two virally encoded proteins, the phosphoprotein (P) and the large protein (L) (Hamagushi *et al.,* 1983; Portner *et al.,* 1988). The L protein contains the polymerase active site, whereas P acts as a cofactor that serves both to stabilize the L protein (Curran *et al.,* 1994; Smallwood *et al.,* 1994) and to place the polymerase complex on the N:RNA template because L alone is unable to interact with the N:RNA (Mellon and Emerson, 1978; Horikami and Moyer 1995). When combined with purified N:RNA, the P–L polymerase can carry out mRNA synthesis *in vitro*. This reaction is strongly stimulated by additional P protein (Curran, 1996; Bowman *et al.,* 1999). So independent of stable complex formation with L, P also plays a supplemental role in RNA synthesis that involves its binding to the N:RNA template. The SeV P protein also forms a complex with unassembled N protein (N°). This complex is the active form of N° protein used to assemble the nascent RNA chain during genome replication (Horikami *et al.,* 1992). The formation of the P–N° complex also prevents the nonspecific aggregation of N° (Curran *et al.,* 1995b).

SeV P protein (568 amino acid) expressed in bacteria or mammalian cells is found as a homooligomer, and deletion analysis has identified a region important for oligomerization (residues 344–411) that is predicted to form a coiled-coil (Curran *et al.,* 1995a). Just near this domain is a region (residues 412–445) whose deletion eliminates P–L complex formation (Curran *et al.,* 1994; Smallwood *et al.,* 1994), and a predicted triple α-helical bundle at the C terminus (residues 479–568) is thought to bind to the N protein of the N:RNA template (Ryan *et al.,* 1991; Curran *et al.,* 1995a; Curran, 1998). The C-terminal 40% of P (residues 344–568) by itself is competent for mRNA synthesis *in vitro* but is inactive for replication (Curran, 1996). Replication requires additionally a short region near the N terminus (residues 33–41) referred to as the chaperone domain because its presence prevents the aggregation of the N protein. Deletion of this domain compromises P–N° complex formation (Curran *et al.,* 1995b).

The oligomeric nature of P is central to its various functions. For example, P binding to the N:RNA involves

[1] To whom reprint requests should be addressed. Fax: 00.33.4.76.88.25.42. E-mail: burmeister@esrf.fr.

0042-6822/00 $35.00

a simultaneous interaction between multiple C-terminal arms of the P oligomer and the exposed C-terminal tail of the assembled N (Curran, 1998). The P protein is envisaged to cartwheel on the template via the simultaneous breaking and reforming of contacts with N subunits such that at least two arms of the oligomer continuously engage the N:RNA, thereby assuring processivity. It is presumed that the P–L polymerase complex binds to the N:RNA similarly. The P–L complex contains only a single catalytic L protein even though there are multiple presumably identical L binding sites on the P oligomer (Curran, 1996). Given the proximity of the L binding site to the oligomerization domain, steric hindrance of the first L bound to P may prevent further L binding. The P proteins of SeV, mumps virus, and Newcastle disease virus (NDV) were suggested to be trimers based on the results of an epitope dilution assay (Curran *et al.,* 1995a; Simonet and Curran, manuscript in preparation).

As part of our on-going structural studies on the SeV P protein, we have overexpressed the C-terminal part of P in bacteria. Partial proteolysis revealed a protease-resistant core that contained the oligomerization domain plus the L binding domain. In this paper, we study the oligomerization of this core by different biochemical and biophysical methods that show that it forms an homotetramer in solution with a very elongated form. The very C-terminal part of P, the X-protein that is expressed in infected cells and that carries the N:RNA binding activity of P (Curran, 1998), is monomeric when expressed alone in *Escherichia coli*.

RESULTS

The sequences of six paramyxovirus P proteins were aligned with the Clustal W program, and the secondary structure was predicted using a threading algorithm (Rost, 1996). The sequences of the N-terminal halves showed very little homology or similarity in predicted secondary structure (not shown). Figure 1 shows the alignment and predictions for the C-terminal half of P (PCT). Similarity in secondary structure prediction was mainly observed for the previously recognized multimerization sequence (Curran *et al.,* 1995a) and for part of the N:RNA binding domain (PX), which is expressed in SeV-infected cells by a mechanism of discontinuous ribosomal scanning of the P mRNA (Curran and Kolakofsky, 1987, 1988). These two regions of the protein are predicted to contain α-helices. It is interesting to note that the set of six proteins does not show much sequence homology but that the secondary structure predictions do show similarities.

To study the structure of SeV P protein we designed three constructs for expression in *E. coli* (Fig. 2). The first construct expresses the C-terminal half of P (PCT) as shown in Fig. 1. This part of the protein supports transcription but not replication as it does not bind to N° (Curran, 1998). The second construct codes for a domain that was defined by limited proteolysis of PCT with trypsin, see Fig. 3. The trypsin resistant peptide of 14.5 kDa was analyzed by mass spectroscopy and N-terminal sequencing. It was found to correspond to a domain beginning with the N terminus of PCT and ending with Arg 446 (see Fig. 1). Because this domain contains the previously identified multimerization sequence, we named it the multimerization domain (PMD). Interestingly, it also includes the complete L binding domain. The PX construct expresses the N:RNA binding domain preceded by a six-residue His tag and a factor Xa cleavage site. All constructs were expressed in bacteria and the resulting proteins were purified under native conditions as described under Materials and Methods. All expressed proteins were soluble and stable, suggesting that they folded as independent folding domains. This was further studied by circular dichroism (CD) measurements.

Figure 4 shows the CD spectra of the recombinant polypeptides. The spectrum of the PMD domain showed a maximum at 193 nm and two minima at 208 and 222 nm. This spectrum is typical for proteins with high α-helical content. Deconvolution of the spectrum using different algorithms suggested a helical content of 52 $\pm$ 7%, in good agreement with the secondary structure prediction in Fig. 1 of 46% helix. Deconvolution of the spectrum for PCT gave 43 $\pm$ 2% helix compared to 33% predicted and for PX deconvolution gave 26 $\pm$ 4% helix compared to 27% predicted. All deconvolutions gave low β-structure contents. Figure 4 also shows a weighted average of the PMD plus the PX spectra that was almost identical to the spectrum for PCT. This confirms that PMD and PX fold as independent domains.

The multimerization domain has a high helical content, in agreement with the prediction that multimerization involves the formation of coiled-coils (Curran *et al.,* 1995a). Other studies have shown that coiled-coils can form dimers, trimers, or tetramers (Harbury *et al.,* 1993, 1994). During the purification procedure, the proteins did not elute from the gel filtration columns as expected from their relative molecular weights. The PCT domain (28.5 kDa) eluted as a 230- to 250-kDa globular protein, corresponding either to a high degree of multimerization or to a very elongated shape of the protein or to both. The PMD domain (14.5 kDa) eluted as a 60-kDa globular protein. The PX domain eluted as a 20-kDa protein, which is too small to be a dimer but too big for a monomer (12.5 kDa). This could be explained either by a monomer with a relatively elongated shape or by a poorly or partly structured protein. Unusual hydrodynamic properties have also been shown for the P protein from vesicular stomatitis virus (VSV) (Gao and Lenard, 1995; Gao *et al.,* 1996).

The global shapes of the P protein domains and their weights were studied by small angle scattering experiments with X-rays (SAXS) and neutrons (SANS). Using

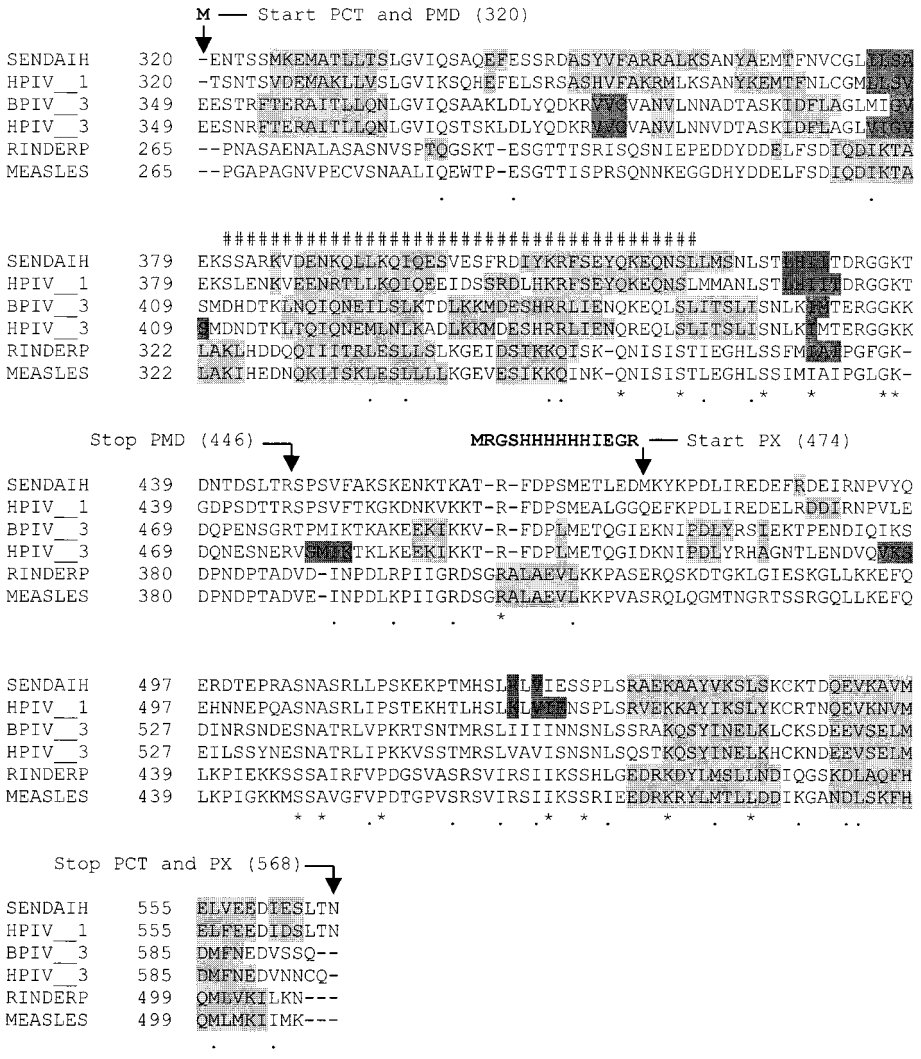

FIG. 1. Multiple alignment and secondary structure prediction for the C-terminal parts of six paramyxovirus P proteins. The multiple alignment was done with Clustal W 1.6 and the secondary structure prediction with the PHD program (Rost, 1996). Stars show strict homology between the six sequences, and dots show the conservative changes. For the secondary structure, α-helices are highlighted in light gray and β-sheets in dark gray. The putative coiled-coil domain is indicated by #, and the start and stop position of PMD and PX are showed by arrows. In bold above the sequence is the Met residue added to the N terminus of PCT and PMD and the poly-histidine tail plus factor Xa cleavage site added to the N terminus of PX. The full names of the viruses are given under Materials and Methods.

X-ray scattering, we determined the global radius of gyration (R_g) of the proteins and the radius of gyration of the cross section (R_c). An example of the data are shown for PMD in Fig. 5A. The derived values for the radii of gyration are 48 Å for PCT and 37 Å for PMD. A plot for the determination of R_c is shown in Fig. 5B. The calculated values for R_c are 14 Å for PCT and 13 Å for PMD. Determination of the weight of the molecular species from SAXS has been calibrated against a particle with known molecular weight. For this we performed a scattering experiment under identical conditions on catalase that has a molecular weight of 240 kDa. Using this calibration, we calculated a weight for PCT of 81 kDa, which corresponds to the weight of a trimer (3 × 28.5 kDa). The weight for PMD was 62 kDa, suggesting that it is a tetramer (4 × 14.5 kDa). Together, the values for R_g,

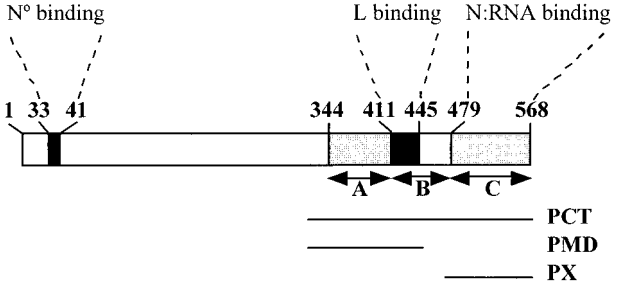

FIG. 2. Schematic representation of the Sendai virus P protein. The P protein is shown as a rectangle, with the various domains indicated by shading. Box A is responsible for the oligomerization of the protein and box C is directly involved in the binding to N:RNA. The lines below the rectangle represent the different parts of the protein studied in this article: PCT, that contains the three boxes A, B, C, PMD that contains box A and the L binding domain and PX that contains box C (after Curran *et al.,* 1995a).

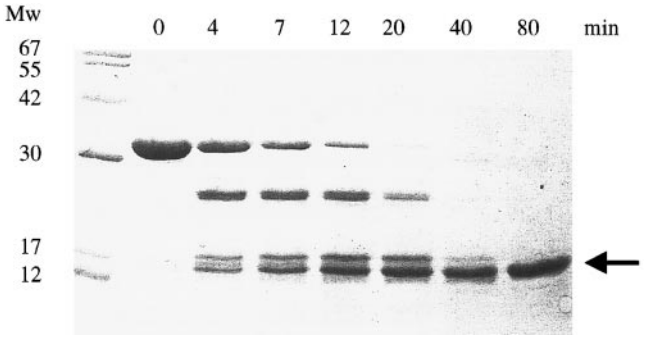

FIG. 3. Limited proteolysis of PCT; 8 μg of PCT was digested with trypsin for different times at room temperature with a PCT/trypsin ratio of 1/500. The reaction was stopped by adding 100 μM of PMSF. The most stable fragment is indicated by an arrow and was analyzed by mass spectrometry and N-terminal sequencing.

R_c, and the derived values for the masses from SAXS suggest that PCT and PMD are trimers or tetramers that are thin and long.

We also performed SANS on PCT (not shown). The data were analyzed in the same way as the data from SAXS and a value for R_g from SANS was derived to be 50 $\pm$ 5 Å. The molecular weight of PCT derived from SANS was 83 kDa but with a significant error of some 20% because the molecular weight determination depends directly on the exact concentration of the protein which is not known to within 10%, adding to the overall error.

To determine whether SeV P is a trimer or a tetramer, we did cross-linking experiments on PMD using two

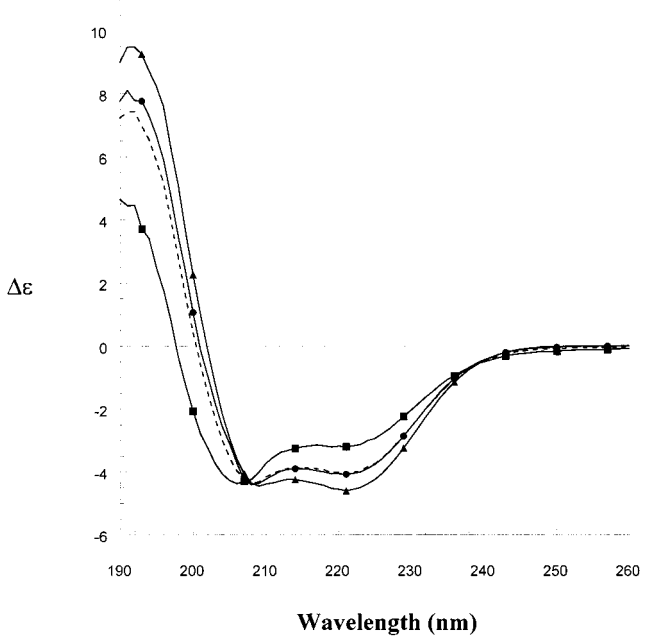

FIG. 4. CD spectra of PCT, PMD, and PX. Far UV CD spectra of PCT (circle), PMD (triangle), and PX (square). The dotted curve shows the weighted average of the PMD and the PX spectra.

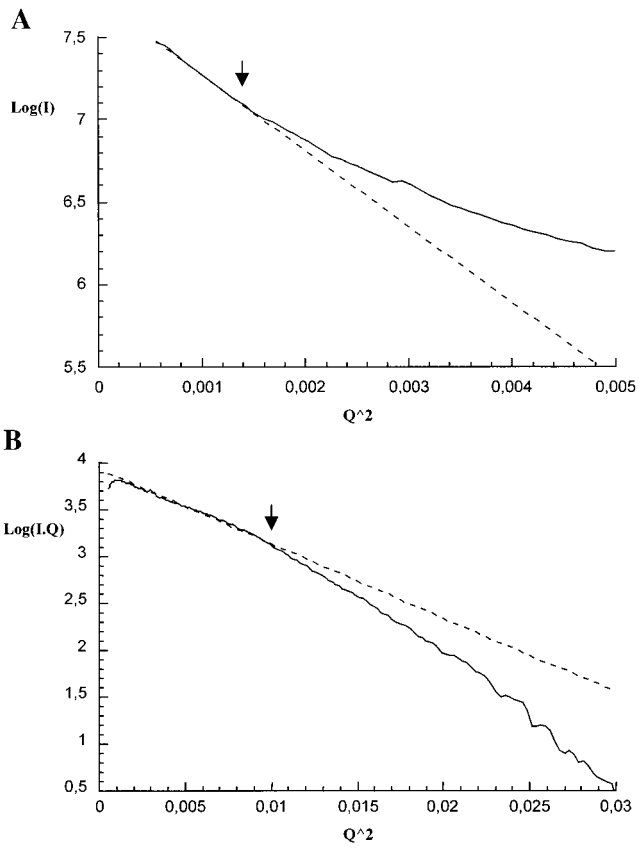

FIG. 5. SAXS plots for PMD. (A) Guinier plot (Log(*I*) *vs.* Q^2) of SAXS data for PMD for the calculation of the radius of gyration. (B) Plot for the calculation of the cross-sectional radius of gyration of PMD (Log(*I.Q*) *vs.* Q^2). The dotted lines are the initial slopes of the curves. Q^2 is expressed in $Å^{-2}$. The maximal limit of the Guinier approximation ($Q \times R = 1.5$) is shown by arrows.

different chemical reagents. Glutaraldehyde is a short (6 Å) self-polymerizing reagent, reacting with lysine, tyrosine, histidine, and tryptophan. SAB is a bifunctional reagent of fixed size (13.1 Å) that reacts only with lysine. The results of the cross-linking experiments on PX showed that this peptide is a monomer (not shown) as expected from its behavior on gel exclusion chromatography. Cross-linking of PMD with glutaraldehyde showed first the formation of tetramers and then of higher molecular weight species that probably consisted of two, three, and four tetramers (Fig. 6A). Cross-linking with SAB showed only formation of tetramers (Fig. 6B). This difference in results is probably caused by the fact that glutaraldehyde forms polymers of various sizes, allowing cross-linking over large distances. Similar cross-linking experiments were done on PCT, but here the results were more difficult to interpret due to aberrant running of the cross-linked species (not shown).

Sedimentation velocity experiments performed on PMD showed a homogeneous species (Fig. 7A). Direct fitting of the boundary profiles allowed to derive, at the three concentrations used, a sedimentation coefficient

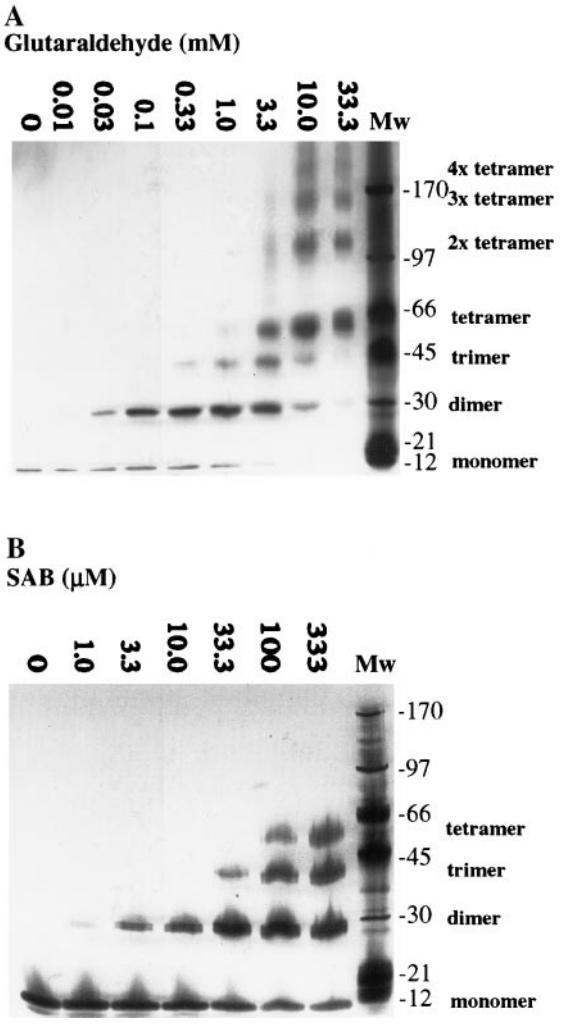

FIG. 6. Chemical cross-linking experiments on PMD. Silver-stained SDS–PAGE showing the results of cross-linking experiments on PMD with glutaraldehyde (A) and SAB (B). The concentrations of glutaraldehyde and of SAB are indicated above (A) and (B), respectively. For conditions, see Materials and Methods.

$S_{20,w}$ of 3.1 S (Fig. 7A) and a molar mass of 57 kDa (expected weight for a tetramer is 58 kDa). Sedimentation equilibrium experiments on PMD were performed for four protein concentrations at three rotor speeds and globally fitted (Fig. 7C) as a homogenous species of 56 kDa, confirming the tetrameric state of PMD. The value of the sedimentation coefficient is consistent with an elongated particle. The sedimentation velocity curves of PCT (Fig. 7B) showed, in addition to a major species of 4.1 S, heavier species (the curves at high radius are not horizontal). The amount of the heavier species was increased at low protein concentrations. SDS–PAGE analysis of the samples after the experiments showed that the protein was slightly degraded, suggesting that the cleaved protein aggregates. Considering two species for the analysis of the sedimentation profiles, we derived a molecular weight for the major species in the PCT preparation of 105 ± 3 kDa (average for the three loading concentrations). It is known that the presence of heterogenieties may lead to an increased widening of the front and to an overestimation of the diffusion coefficient and thus to an underestimation of the molecular weight. Therefore, we independently determined a hydrodynamic radius (*Rh*) of PCT by gel filtration of 60 Å (using a Boehringer calibration kit: Combithek calibration proteins for range 12.5–450 kDa; *Rh* of 17 to 63 Å). Combining *Rh* and *S* values, we derived a weight for PCT of 107 kDa, which corresponds to 3.8 monomers. The sedimentation equilibrium profiles obtained for PCT could not be fitted properly due to the polydispersity of the sample. The results of the sedimentation experiments suggest that both PMD and PCT are extended tetramers.

We then performed a His-tag dilution experiment in which His-tagged PCT and non-tagged PCT were mixed at various ratios, denatured with 8 M urea, and then renatured by step-wise dialysis to remove the urea. Upon renaturation, mixed oligomers were fished out of the solution with Ni–NTA activated beads and analyzed by SDS–PAGE, an example of which is shown in Fig. 8A. The tagged and nontagged protein bands were well separated. They were quantified with a Shimadzu scanner, and the ratio of untagged to tagged protein was plotted against the input ratio on a graph that also shows the theoretical curves for trimeric, tetrameric, and pentameric proteins (Fig. 8B). These theoretical curves were calculated with the probability formula given under Materials and Methods. As can be seen from Fig. 8B, the experimental points correspond most closely to the theoretical curve for the tetramer.

Finally, we crystallized the PMD domain. The PMD domain crystallizes as thin plates of 200 × 300 × 5 μm. The crystals diffract up to 3.5 Å resolution. They belong to spacegroup $C222_1$ with cell parameters of $a = 48.7$ Å, $b = 498$ Å, $c = 48.7$ Å, $\alpha = \beta = \gamma = 90°$ and are systematically twinned.

DISCUSSION

The results presented in this paper address three aspects of the structure of the C-terminal part of the SeV P protein: its domain structure, its multimerization state, and its overall physical shape and size.

Domain structure

The C-terminal part of P can fold without the N-terminal part into an active protein (Curran, 1998) that is soluble and that has the secondary structure content predicted from its sequence. The C-terminal part consists of two subdomains, the multimerization domain, including the L-binding site, and the X-protein that both also fold independently into stable units. The C-terminal end of the multimerization domain was defined by limited proteolysis. The N terminus of PX is defined as the N terminus of the protein that is produced in the infected

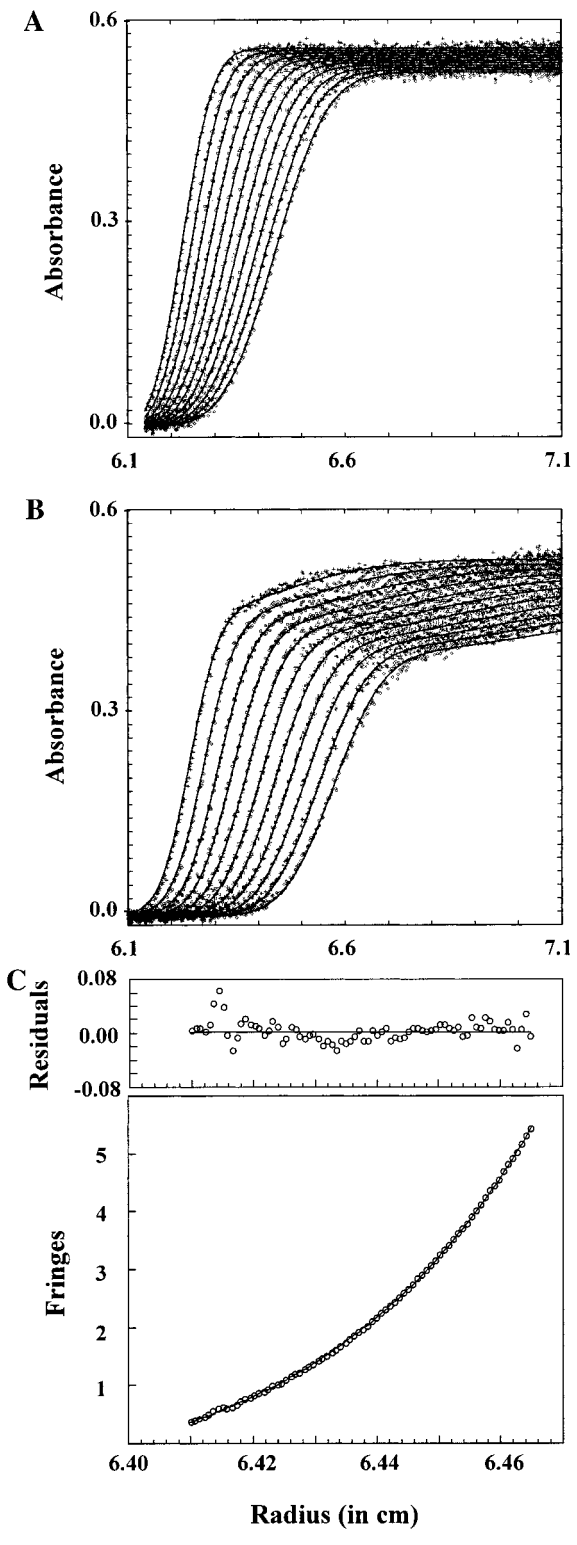

FIG. 7. Sedimentation experiments with PMD and PCT. (A and B) Measurements of sedimentation velocity. An overlay of profiles recorded approximately 11 min apart during 2 h is shown. The fitting curves resulting from the analysis by Svedberg software are shown in solid lines, and the data in each curve are shown with different symbols. (A) Experiment on PMD, the fitting model is with one species, resulting in an *S* value of 3.1 S and a derived Mw of 57 kDa. (B) Experiment on PCT, the fitting model uses two species. The main species has an *S* value of 4.1 S. (C) Equilibrium sedimentation measurement on PMD. The data are presented by open circles and the fitted curves by a solid line. This plot is for an initial concentration of 0.75 mg/ml and a speed of 25,000 rpm, the residuals are plotted above. The three highest concentrations and the three speeds were fitted simultaneously yielding a Mw of 56 kDa.

cell. The two subdomains are connected by a rather variable sequence without strongly predicted secondary structure. Because PX was found to be monomeric, this connecting sequence possibly connects the monomeric X to the oligomeric rest of the C-terminal part of P.

Multimerization state

The present results on the cross-linking of the multimerization domain, the sedimentation analysis on PMD and PCT, and the His-tag dilution experiment on PCT suggest that SeV P forms a tetramer. However, the P protein of SeV is generally thought to be a trimer. The results of early work by Markwell and Fox (1980), who performed reversible cross-linking studies and two-dimensional PAGE analysis on virus, were interpreted as P being a trimer. However, their Fig. 5 shows thick bands for the dimer and tetramer of HN at exactly the positions where dimers and tetramers of P could have been expected. Furthermore the supposed trimer migrates as a protein with a molecular weight of 240 kDa, whereas the monomer weighs 62 kDa. Therefore in our opinion, their work does not exclude that P is a tetramer. Curran *et al.* (1995a) used an epitope dilution assay and showed that, even when there was 7 or 15 times more of the untagged protein co-expressed than the tagged protein that was pulled down with an antibody, the ratio between untagged and tagged protein in the immunoprecipitate was only 2. From this they concluded that the multimer must be a trimer. However, at an input ratio of 7 times more untagged over tagged, from the probability formula given under Methods and Materials, one may derive that the theoretical output ratio should have been 1.64 untagged/tagged for a trimer or 2.31 for a tetramer. Similar experiments on the P proteins of Mumps Virus and NDV resulted in ratios of 1.6 to 1.7 untagged over tagged (Curran *et al.,* 1995a; Simonet and Curran, unpublished results).

There are a number of possible reasons why the results and their interpretation from the previously published epitope dilution assays differ from the present results. First, even at considerable untagged/tagged input ratios, there are still multimers that contain more than one tagged monomer. This is the reason why at an input ratio of 7 untagged/tagged the output ratio is only 1.64 for a trimer and not 2. Second, the method used by Curran *et al.* (1995a) uses co-expression of untagged and tagged protein in transfected eukaryotic cells. Even at high untagged/tagged ratios, there will always be a certain sta-

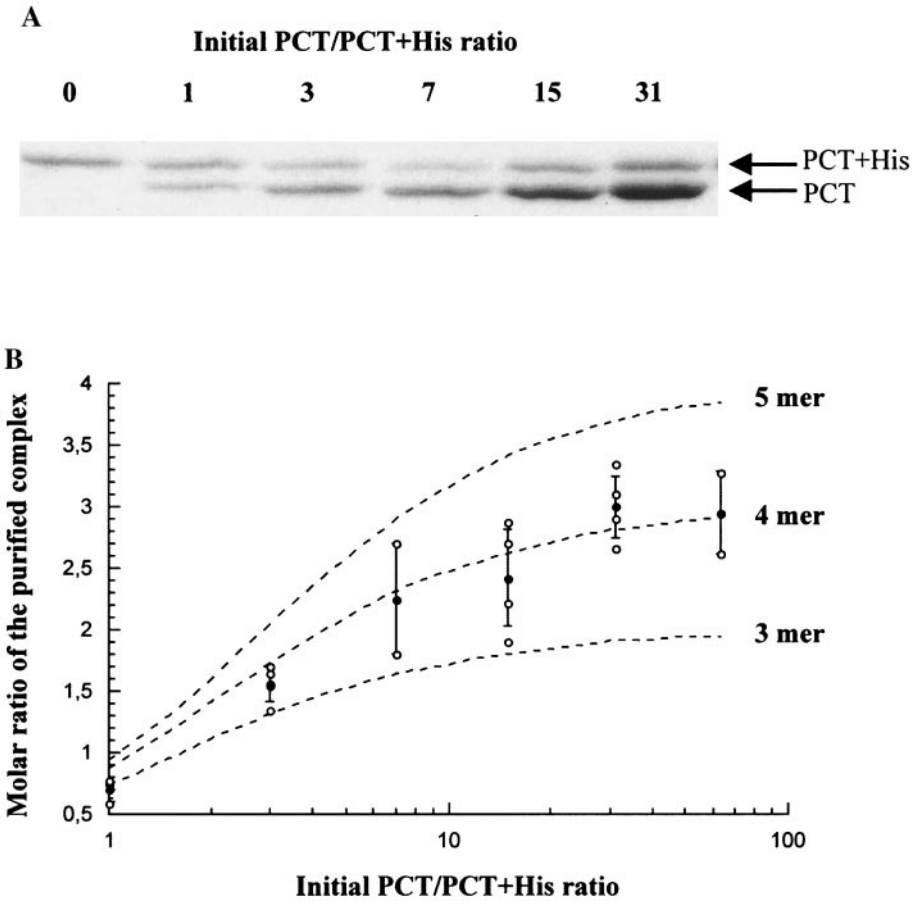

FIG. 8. His-tag dilution experiment. Untagged and tagged PCT were mixed at different ratios and then denatured in 8M urea and renatured by step dialysis. Tagged multimers were recovered from the mixture using Ni–NTA-activated beads. Bound proteins were separated on SDS–PAGE, stained with Coomassie blue and quantitated by densitometry. (A) Typical migration of mixed oligomers on SDS–PAGE after recovery of the nickel-bound proteins. The initial mixing ratios are indicated above the gel. (B) Plot of the ratio PCT/PCT+His recovered from the nickel beads as a function of the initial mixing ratio. Dotted lines are the theoretical curves for trimer, tetramer and pentamer. Open circles are the experimental measurements and black circles are the mean values with standard deviation.

tistic distribution of the two plasmids used for transfection. Because the dilution of the tag is revealed by fishing out only the multimers that contain at least one tag, the statistical variation in the transfection will lead to a higher representation of the tagged form of the protein. Third, in the tag dilution assay using cotransfection there may exist cotranslational multimerization on the polysomes, as described for the trimerization of the reovirus cell-attachment protein (Gilmore *et al.,* 1996), which would also lead to an overrepresentation of the tagged protein and to the conclusion of a lower-number multimer. For a tag-dilution assay to be valid, there needs to be free exchange between the monomers and the actual input ratio at the moment of multimerization needs to be known. Such conditions were used by Gao *et al.* (1996), who performed a His-tag dilution assay with VSV P protein. Although the authors concluded that VSV P protein was a trimer, the results published in their Fig. 1 correspond most closely to the theoretical curve of the tetramer shown in our Fig. 8B. However, from this to conclude that VSV P protein is also a tetramer would be preliminary since the highest untagged/tagged input ratio they used was only 8, whereas the highest ratio we used here was 63.

In vivo, the SeV P protein is phosphorylated but the proteins that we have worked with in the present study are expressed in *E. coli* and are not phosphorylated. The constitutively phosphorylated residue in the SeV P protein, serine 249, (Byrappa *et al.,* 1996) lies in the N-terminal half of the protein and is not present in the constructs examined in this study. Therefore phosphorylation is unlikely to play a role in modulating the multimerization state of the SeV P protein. This contrasts with the situation observed with VSV in which phosphorylation induces oligomerization of P and it is the oligomeric form of P that interacts with L and supports transcription (Gao and Lenard, 1995; Gao *et al.,* 1996; Spadafora *et al.,* 1996). The form of VSV P that supports replication is bound to N° with a 1:1 stoichiometry (Masters and Banerjee, 1988; Peluso and Moyer, 1988), and

recent unpublished results suggest that this complex consists of 1 N plus 1 P molecule (Iseni and Ruigrok, unpublished). It is possible that P in the N°-P complex is not phosphorylated, which would open up the possibility that the phosphorylation status of P may play a pivotal role in regulating the relative rates of transcription and replication.

Overall physical shape and size

If we assume that PMD is a long rod then we can calculate the dimensions of the rod using: $R_g^2 - R_c^2 = L^2/12$ where L is the length of the rod and $R_c^2 = (D/2)^2/2$ where D is the diameter of the particle. For PMD this would give a length of 121 Å and a diameter of 36 Å. The structure of the coiled-coil part of P could look like the structure of the SNARE 4 helix coiled-coil (Sutton *et al.,* 1998), which has a diameter of 23 Å. This information plus the fact that a cylindrical particle with a length of 121 Å and a width of 36 Å would be much to big for a tetramer may suggest that PMD has a dumbbell- or bouquet-like shape. The crystals also give some information about the global shape of the multimerization domain of the SeV P protein. The rather unexpected cell parameters (a = 48.7 Å, b = 498 Å, c = 48.7 Å) for a protein of 4 × 14.5 kDa suggest a very elongated shape. The radius of gyration for PCT is even larger than that of PMD, but the two values for R_c are almost the same. Because PX was found to be monomeric, it is possible that the X domains extend more or less flexibly from one end of the PMD molecule, adding to the length and/or diameter of PMD. Such a structure could correspond to a protein that cartwheels over the N:RNA template. The fact that the P protein is a tetramer rather than a trimer has no consequence for the models that have been proposed for P function on the template (Curran *et al.,* 1995a). The cartwheeling model requires contact between at least two (but not one) arms of the P oligomer and the nucleocapsid to ensure stability, and the simultaneous breaking and reforming of these contacts to ensure processivity as the P or P–L complex move on the template. This can be realized with either a trimer or a tetramer but not with a dimer.

MATERIALS AND METHODS

Constructs and protein purification

All constructs were derived from an initial pT7–7 plasmid containing the full-size P gene from the Sendai Harris strain with a 6 × His tag (Curran, 1998). The different constructs described below were obtained by PCR-directed mutagenesis using specific oligonucleotides (Genosys) and Pfu polymerase (Stratagene). The different proteins were overexpressed in *E. coli* BL21 (DE3) by induction with 0.2 mM IPTG as the culture reached an OD_{600} of 0.6.

C-terminal domain of P (PCT)

PCT (residues 320–568, 28.5 kDa) was overexpressed at 23°C for 6 h. Cells were lysed by sonication in 50 mM Tris–HCl, pH 8; 50 mM NaCl; and 10 mM EDTA in the presence of Pefabloc (Sigma) and DNase I (Boehringer). Cell debris was eliminated by centrifugation at 23,000 *g* for 15 min. The filtrate through a 0.45-μm pore filter was loaded on a Q Sepharose Fast Flow (Pharmacia) 5- ml column. The protein was eluted with an NaCl gradient (50–500 mM) in 50 mM Tris, pH 8. Precipitation of the fractions that contained the protein with 40% ammonium sulfate on ice was followed by centrifugation at 10,000 *g* for 10 min and resuspension in 10 mM sodium phosphate, pH 6.8. The resuspended protein was run on a hydroxyapatite CHT2-I column (Bio-Rad) using the Biologic HR system (Bio-Rad) and eluted with a sodium phosphate gradient (10–500 mM). The protein solution was then reduced to 2 ml by centrifugation using a Centricon 50 (Amicon), loaded on a Superdex 200 Hiload 16/60 column (Pharmacia) equilibrated in 20 mM HEPES, pH 7.7, and 200 mM NaCl, and finally concentrated by Centricon centrifugation.

C-terminal domain of P with His tag (PCT+His)

PCT+His (MRGSHHHHHHTNSRPG + residues 320–568, 30.2 kDa) was overexpressed under the same conditions as PCT. Cells were lysed by sonication and the protein purified on a Ni–NTA superflow (Quiagen) column according to the manufacturer's protocol. The eluted protein was then reduced by concentration to 2 ml and loaded on a Superdex 200 Hiload 16/60 column as described for PCT.

Multimerization domain of P (PMD)

PMD (residues 320–446, 14.5 kDa) was overexpressed at 30°C for 4 h. Cells were lysed by sonication in 50 mM HEPES, pH 7.0, and 10 mM EDTA in the presence of Pefabloc and DNase I. The soluble extract was prepared as described above and then loaded on a SP Sepharose fast Flow (Pharmacia) 5-ml column. The protein was eluted with an NaCl gradient (0–500 mM) in 50 mM HEPES, pH 7.0. Precipitation of the fractions in 50–100 mM with 50% ammonium sulfate was followed by centrifugation at 10,000 *g* for 15 min. PMD was then resuspended in 10 mM potassium phosphate, pH 6.8 and loaded on a hydroxyapatite column as for the PCT fragment. The volume was then reduced to 2 ml by centrifugation using Centricon 30 and the protein loaded on a Superdex 75 FPLC column (Pharmacia) equilibrated in 20 mM HEPES, pH 7.0; 50 mM NaCl; and 1 mM DTT and finally concentrated on Centricon 30.

X domain of P (PX)

The X-protein (residues 474–568, 12.5 kDa) with a 6 × His tag and a factor Xa cleavage site was overexpressed

at 37°C for 4 h. Cells were lysed by sonication in 40 mM Tris–HCl, pH 8.0; 500 mM NaCl; and 5 mM imidazole in the presence of Pefabloc and DNase I. The soluble extract was then loaded on a 3-ml column filled with Sepharose CL 6-B (Pharmacia) activated with nickel. Elution with the same buffer plus 1 M imidazole was followed by the addition of 20 mM EDTA and 1.5 M ammonium sulfate. The PX containing fractions were then loaded on a Phenyl-superose FPLC column (Pharmacia) equilibrated in 50 mM Tris–HCl, pH 8.0; 200 mM NaCl; and 1.5 M ammonium sulfate, and elution was performed with an ammonium sulfate inverted gradient (1.5–0 M). The protein was concentrated by centrifugation using Centricon 10 before loading on a Superdex 75 FPLC column equilibrated in 20 mM HEPES, pH 7.7, and 200 mM NaCl and then further concentrated.

Determination of the protein concentration

The absorption coefficients were calculated for each fragment with the formula: ϵ = 1480 × (number of Tyr)/Mw. The resulting $\epsilon_{280nm,\ 1cm}$ were 0.36 for PCT, 0.41 for PMD, and 0.35 for PX. These values are low because the fragments of SeV P contain no tryptophans. The absorption coefficient for PCT was also experimentally determined by determining the OD_{280nm} and the amino acid composition after hydrolysis of a given sample. The absorption coefficient was determined to be 0.32 with an estimated error of between 5 and 10%. For all experiments with PCT, the value of 0.32 was used.

Structure prediction and multiple alignment

We compared the P proteins from six different paramyxoviruses: Sendai virus Harris strain, human parainfluenza virus type 1 and 3, bovine parainfluenza virus type 3, rinderpest virus, and measles virus. The structure predictions were done with the PredictProtein prediction package, (PHD, EMBL, Heidelberg) (Rost, 1996) with normal penalty and by using each sequence separately. The multiple alignment was done with Clustal W 1.6 (Thompson *et al.,* 1994) with a gap opening penalty of 5 and a gap extension penalty of 0.4.

Circular dichroism

CD experiments were done on a JOBIN Yvon CD6 spectrophotometer, with a 1- mm cell. The spectra were collected from 190 to 260 nm with 1- nm steps and an integration time of 5 s/nm. Two spectra were done for each measurement and averaged. The proteins were in 50 mM potassium phosphate, pH 6.8 and 150 mM KF at a concentration of ~0.33 mg/ml. Protein concentrations were determined by OD_{280} measurements, see above. The measurements were corrected for buffer absorption and the resulting spectra expressed as $\Delta\epsilon = \Delta A/C \cdot L$ where ΔA is the corrected measure, C is the molar concentration of the protein multiplied by the number of amino acids in the considered fragment, and L is the thickness of the cuvette. The resulting spectra were deconvoluted by CDnn, K2D, and DicroProt software (Bohm *et al.,* 1992; Andrade *et al.,* 1993). These programs are available on http://bioinformatik.biochemtech.uni-halle.de/cdnn/index.html.

Cross-linking experiments

PMD was incubated with glutaraldehyde for 16 h at 20°C with a fixed amount of protein (5 μg in 30 μl) and different amounts of cross-linker (0–33 mM). The reactions were quenched by adding 100 mM glycine. The samples were analyzed on a 4–15% polyacrylamide denaturing gel and silver stained on a PhastSystem (Pharmacia) according to protocol of the manufacturer. Cross-linking with Suberic Acid Bis(*N*-hydroxy-succinimide ester) (SAB, Sigma) was also done for 16 h at 20°C, using the same amount of protein but 0–0.33 mM of cross-linker and the samples were analyzed on a 10–15% acrylamide gel. The SAB was first solubilised in DMSO at a concentration of 10 mM and then diluted in 50 mM HEPES, pH 7.0, and 150 mM NaCl to the desired concentration.

Small angle X-ray and neutron scattering experiments (SAXS and SANS)

SAXS was performed at the ESRF (Grenoble, France) on the protein crystallography beamline ID14–3 at a wavelength of 0.933 Å using a two-dimensional MarCCD detector. The sample was measured in a cuvette with 5-μm Mylar sheet windows with a 1-mm path length. The sample to detector distance was 520 mm, and a helium path was inserted between sample and detector. We collected scattering curves at two protein concentrations (typically 2 and 5 mg/ml) and a blank containing only buffer. Four measurements were averaged and data were reduced with FIT-2D (Hammersley *et al.,* 1996). The dose for each exposure was measured with an ion chamber and was used as a multiplication factor in the background subtractions. An additional factor in the background subtraction of 0.97 was used to obtain a background of zero outside the resolution range of a measurable small-angle scattering signal.

SANS was performed on the D22 instrument at the Institut Laue-Langevin (Grenoble, France) using a wavelength of 10 Å and a sample-detector distance of 8 m. The protein concentration was 5 mg/ml.

Both methods give information on the shape of the molecular species in solution. The global radius of gyration is derived from the Guinier approximation: $I(Q) = I_o \exp(-R_g^2 Q^2/3)$ where I_o is the zero angle scattered intensity, R_g the overall radius of gyration, and $Q = 4\pi\sin\theta/\lambda$ with 2θ being the scattering angle and λ the wavelength. The radius of gyration can be calculated from the initial slope of the plot Log(I) *vs.* Q^2. We also

calculated the radius of gyration of the cross-section (R_c) assuming elongated rod-like particles (Glatter and Kratky, 1982) from the formula: $Q.I(Q) = I_c\exp(-R_c^2Q^2/2)$. R_c was calculated from the initial slope of the plot $\mathrm{Log}(I.Q)$ *vs.* Q^2.

The mass of the molecular species can also be derived from the scattering experiments. For the neutron scattering experiment, the mass could be directly calculated from I_o using the exact protein concentration (Jacrot and Zaccai, 1981). For the X-ray scattering experiment, the correlation between I_o and the mass needed to be calibrated using a protein with known molecular weight but again depended directly on the protein concentration.

Analytical ultracentrifugation

Measurements of the velocity sedimentation and equilibrium sedimentation were performed on PCT and PMD proteins with a Beckman XL-I analytical ultracentrifuge and an AN-60 rotor (Beckman instruments). For the measurements of the velocity sedimentation, we used three concentrations of 0.5, 1.0, and 1.5 mg/ml in 20 mM HEPES, pH 7.7; 200 mM NaCl; and 0.02% NaN_3 in double-sector cells with epon centerpieces and sapphire windows. The experiments were done at 42,000 rpm, 20°C, and data were collected every 320 s by absorbance measurements at 277 nm for a duration of ≥3 h. Equilibrium sedimentation measurements on PMD were done with four protein concentrations (0.37; 0.75; 1.5, and 3 mg/ml) using eight hole epon centerpieces with sapphire windows. The measurements were done at 14,000, 18,000, and 25,000 rpm, 4°C during 5 h for each speed. The data were collected every 20 min in the interference mode. The data were analyzed by Svedberg software (V6.23; written by J. Philo; http://www.jphilo.mailway.com) for the velocity experiments and the Beckman software for the equilibrium. The partial specific volume, solvent density and viscosity were calculated with Sednterp software (V1.01 from D.B. Haynes, T. Lane and J. Philo; http:///www.bbri.harvard.edu/rasmb/rasmb.html).

His-tag dilution assay

The PCT and PCT+His proteins were mixed in different ratios (from 1:1 up to 63 :1, PCT:PCT+His) and then denatured in 8 M urea; 20 mM Tris–HCl, pH 8.0; 200 mM NaCl; 2.5 mM EDTA; and 0.5% N-P40. For all ratios, the total protein concentration was the same (0.16 mg/ml) to maintain the same renaturation conditions. The experiment has also been performed with a four times higher protein concentration with the same results. Then the mixed and denatured proteins were dialyzed at 4°C in Slide-A-Lyser cassette (Pierce) cutoff 10 kDa against the same buffer with successively 6, 4, 3, 2, 1, and 0.5 M urea for 1 h per step. Two more steps were done in 20 mM Tris–HCl, pH 8.0, and 200 mM NaCl. The renaturated proteins were mixed with 100 μl of Ni–NTA-activated beads (Quiagen) for 100 μg protein at 4°C for 30 min. The beads where then spun down at 2000 rpm and washed three times with 20 volumes of buffer. The beads with the attached proteins were then mixed with one volume of 2× SDS–PAGE loading buffer containing 100 mM EDTA and directly loaded onto a 15% denaturing gel. The resulting bands were stained by Coomassie R250 blue (Pharmacia) and quantified with a Shimadzu CS-9000 gel scanner. The theoretical ratio (R = PCT/PCT + His) resulting from the random renaturation and the selection of the oligomers that contained at least one tagged monomer is described by the formula

$$R = \frac{\sum_{i=1}^{n}(n-i)C_n^i p^i (1-p)^{n-i}}{\sum_{i=1}^{n} iC_n^i p^i (1-p)^{n-i}}$$

Where p is the input fraction of tagged protein (PCT+His/total protein) and n is the number of monomers in the oligomer. The mathematical derivation of this formula was done by Dr. Yves Gaudin (CNRS, Gif-sur-Yvette, France) and an explanation of this formula may be obtained from the authors.

Crystallization conditions.

PMD was crystallized by the vapor-diffusion hanging-drop method. We used a reservoir solution of 100 mM glycine, pH 10.0; 200 mM NaCl; 10 mM $CaCl_2$; 14% PEG 2000 mono methyl ether; and 9% PEG 400. The protein was in a buffer containing 20 mM HEPES, pH 7.0;, 200 mM NaCl; and 1 mM DTT at a concentration of 5 mg/ml. Crystals appeared after 3–5 days at 20°C.

ACKNOWLEDGMENTS

We thank Mathias Wilm (EMBL, Heidelberg) for mass spectroscopy, Jean Gagnon (IBS, Grenoble) for amino acid analysis, the Institut Laue-Langevin for beam-time on D22, and Peter Timmins (ILL, Grenoble) for help with data collection and interpretation, Dominique Madern (IBS, Grenoble) for technical assistance during the CD spectra acquisition, Jean-Baptist Marq (CMU, Geneva) for technical assistance, and Yves Gaudin (CNRS, Gif-sur-Yvette) for discussions and for the derivation of the His-tag dilution probability formula.

REFERENCES

Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.* **6,** 383–390.

Bohm, G., Muhr, R., and Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng.* **5,** 191–195.

Bowman, M. C., Smallwood, S., amd Moyer, S. A. (1999). Dissection of individual functions of the Sendai virus phosphoprotein in transcription. *J. Virol.* **73,** 6474–6483.

Byrappa, S., Pan, Y. B., and Gupta, K. (1996). Sendai virus P protein is constitutively phosphorylated at serine 249: High phosphorylation potential of the P protein. *Virology* **216,** 228–234.

Calain, P., and Roux, L. (1993). The rule of six, a basic feature for

efficient replication of Sendai virus defective interfering RNA. *J. Virol.* **67,** 4822–4830.

Curran, J. (1996). Reexamination of the Sendai virus P protein domains required for RNA synthesis: A possible supplemental role for the P protein. *Virology* **221,** 130–140.

Curran, J. (1998). A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* **72,** 4274–80.

Curran, J., Boeck, R., Lin-Marq, N., Lupas, A., and Kolakofsky, D. (1995a). Paramyxovirus phosphoproteins form homotrimers as determined by an epitope dilution assay, via predicted coiled coils. *Virology* **214,** 139–149.

Curran, J. A., and Kolakofsky, D. (1987). Identification of an additional Sendai virus non-structural protein encoded by the P/C mRNA. *J. Gen. Virol.* **68,** 2515–2519.

Curran, J., and Kolakofsky, D. (1988). Scanning independent ribosomal initiation of the Sendai virus X protein. *EMBO J.* **7,** 2869–2874.

Curran, J., Marq, J. B., and Kolakofsky, D. (1995b). An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J. Virol.* **69,** 849–855.

Curran, J., Pelet, T., and Kolakofsky, D. (1994). An acidic activation-like domain of the Sendai virus P protein is required for RNA synthesis and encapsidation. *Virology* **202,** 875–884.

Egelman, E. H., Wu, S. S., Amrein, M., Portner, A., and Murti, G. (1989). The Sendai virus nucleocapsid exists in at least four different helical states. *J. Virol.* **63,** 2233–2243.

Gao, Y., Greenfield, N. J., Cleverley, D. Z., and Lenard, J. (1996). The transcriptional form of the phosphoprotein of vesicular stomatitis virus is a trimer: Structure and stability. *Biochemistry* **35,** 14569–14573.

Gao, Y., and Lenard, J. (1995). Multimerization and transcriptional activation of the phosphoprotein (P) of vesicular stomatitis virus by casein kinase-II. *EMBO J.* **14,** 1240–1247.

Gilmore, R., Coffey, M. C., Leone, G., McLure, K., and Lee, P. W. K. (1996). Co-translational trimerization of the reovirus cell attachment protein. *EMBO J.* **15,** 2651–2658

Glatter, O., and Kratky, O. (1982). "Small Angle X-ray Scattering." Academic Press. New York.

Hamaguchi, M., Yoshida, T., Nishikawa, K., Naruse, H., and Nagai, Y. (1983). Transcriptive complex of Newcastle disease virus. Both L and P proteins are required to constitute an active complex. *Virology* **128,** 105–117.

Hammersley, A. P., Svensson, S. O., Hanfland, M., Fitch, A. N., and Hausermann, D. (1996). Two dimensional detector software: From real detector to idealised image or two theta scan. *High Pressure Res.* **14,** 235–248.

Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262,** 1401–1407.

Harbury, P. B., Kim, P. S., and Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature* **371,** 80–83.

Horikami, S. M., Curran, J., Kolakofsky, D., and Moyer, S. A. (1992). Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro. *J. Virol.* **66,** 4901–4908.

Horikami, S. M., and Moyer, S. A. (1995). Alternative amino acids at a single site in the Sendai virus L protein produce multiple defects in RNA synthesis in vitro. *Virology* **211,** 577–582.

Jacrot, B., and Zaccai, G. (1981). Determination of molecular weight by neutron scattering. *Biopolymers* **20,** 2413–2426.

Kolakofsky, D., Pelet, T., Garcin, D., Hausmann, S., Curran, J., and Roux, L. (1998). Paramyxovirus RNA synthesis and the requirement for hexamer genome length: The rule of six revisited. *J Virol* **72,** 891–899.

Markwell, M. A., and Fox, C. F. (1980). Protein–protein interactions within paramyxoviruses identified by native disulfide bonding or reversible chemical cross-linking. *J. Virol.* **33,** 152–166.

Masters, P. S., and Banerjee, A. K. (1988). Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus. *J. Virol.* **62,** 2651–2657.

Mellon, M. G., and Emerson, S. U. (1978). Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. *J. Virol.* **27,** 560–567.

Peluso, R. W., and Moyer, S. A. (1988). Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* **162,** 369–376.

Portner, A., Murti, K. G., Morgan, E. M., and Kingsbury, D. W. (1988). Antibodies against Sendai virus L protein: Distribution of the protein in nucleocapsids revealed by immunoelectron microscopy. *Virology* **163**, 236–239.

Rost, B. (1996). PHD: Predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.* **266,** 525–539.

Ryan, K. W., Morgan, E. M., and Portner, A. (1991). Two noncontiguous regions of Sendai virus P protein combine to form a single nucleocapsid binding domain. *Virology* **180,** 126–134.

Smallwood, S., Ryan, K. W., and Moyer, S. A. (1994). Deletion analysis defines a carboxyl-proximal region of Sendai virus P protein that binds to the polymerase L protein. *Virology* **202,** 154–163.

Spadafora, D., Canter, D. M., Jackson, R. L., and Perrault, J. (1996). Constitutive phosphorylation of the vesicular stomatitis virus P protein modulates polymerase complex formation but is not essential for transcription or replication. *J. Virol.* **70,** 4538–4548.

Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* **395,** 347–353.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.